



# Complexity of the 5' Untranslated Region of *EIF4A3*, a Critical Factor for Craniofacial and Neural Development

Gabriella S. P. Hsia<sup>1†</sup>, Camila M. Musso<sup>1†</sup>, Lucas Alvizi<sup>1</sup>, Luciano A. Brito<sup>1</sup>, Gerson S. Kobayashi<sup>1</sup>, Rita C. M. Pavanello<sup>1</sup>, Mayana Zatz<sup>1</sup>, Alice Gardham<sup>2</sup>, Emma Wakeling<sup>2</sup>, Roseli M. Zechi-Ceide<sup>3</sup>, Debora Bertola<sup>1,4</sup> and Maria Rita Passos-Bueno<sup>1\*</sup>

<sup>1</sup> Centro de Estudos do Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, <sup>2</sup> North East Thames Genetics Service, Great Ormond Street Hospital, London, United Kingdom, <sup>3</sup> Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, São Paulo, Brazil, <sup>4</sup> Instituto da Criança, Hospital das Clínicas da FMUSP, Universidade de São Paulo, São Paulo, Brazil

## OPEN ACCESS

### Edited by:

Babajan Banganapalli,  
King Abdulaziz University,  
Saudi Arabia

### Reviewed by:

Munis Dundar,  
Erciyes University, Turkey  
Richard G. Boles,  
Center for Neurological  
and Neurodevelopmental Health,  
United States

### \*Correspondence:

Maria Rita Passos-Bueno  
passos@ib.usp.br

<sup>†</sup>Co-first authors

### Specialty section:

This article was submitted to  
Genetic Disorders,  
a section of the journal  
Frontiers in Genetics

**Received:** 15 February 2018

**Accepted:** 10 April 2018

**Published:** 25 April 2018

### Citation:

Hsia GSP, Musso CM, Alvizi L,  
Brito LA, Kobayashi GS,  
Pavanello RCM, Zatz M, Gardham A,  
Wakeling E, Zechi-Ceide RM,  
Bertola D and Passos-Bueno MR  
(2018) Complexity of the 5'  
Untranslated Region of *EIF4A3*,  
a Critical Factor for Craniofacial  
and Neural Development.  
*Front. Genet.* 9:149.  
doi: 10.3389/fgene.2018.00149

Repeats in coding and non-coding regions have increasingly been associated with many human genetic disorders, such as Richieri-Costa-Pereira syndrome (RCPS). RCPS, mostly characterized by midline cleft mandible, Robin sequence and limb defects, is an autosomal-recessive acrofacial dysostosis mainly reported in Brazilian patients. This disorder is caused by decreased levels of *EIF4A3*, mostly due to an increased number of repeats at the *EIF4A3* 5'UTR. *EIF4A3* 5'UTR alleles are CG-rich and vary in size and organization of three types of motifs. An exclusive allelic pattern was identified among affected individuals, in which the CGCA-motif is the most prevalent, herein referred as "disease-associated CGCA-20nt motif." The origin of the pathogenic alleles containing the disease-associated motif, as well as the functional effects of the 5'UTR motifs on *EIF4A3* expression, to date, are entirely unknown. Here, we characterized 43 different *EIF4A3* 5'UTR alleles in a cohort of 380 unaffected individuals. We identified eight heterozygous unaffected individuals harboring the disease-associated CGCA-20nt motif and our haplotype analyses indicate that there are more than one haplotype associated with RCPS. The combined analysis of number, motif organization and haplotypic diversity, as well as the observation of two apparently distinct haplotypes associated with the disease-associated CGCA-20nt motif, suggest that the RCPS alleles might have arisen from independent unequal crossing-over events between ancient alleles at least twice. Moreover, we have shown that the number and sequence of motifs in the 5'UTR region is associated with *EIF4A3* repression, which is not mediated by CpG methylation. In conclusion, this study has shown that the large number of repeats in *EIF4A3* does not represent a dynamic mutation and RCPS can arise in any population harboring alleles with the CGCA-20nt motif. We also provided further evidence that *EIF4A3* 5'UTR is a regulatory region and the size and sequence type of the repeats at 5'UTR may contribute to clinical variability in RCPS.

**Keywords:** acrofacial dysostosis, non-coding region, haplotype, expansion, crossing-over

## INTRODUCTION

Over two-thirds of the human genome is comprised by repetitive elements (de Koning et al., 2011), which have been increasingly associated with functional regulatory roles. Consequently, a variety of human genetic disorders are caused by repeats in coding and non-coding sequences (Cummings and Zoghbi, 2000; Gatchel and Zoghbi, 2005; Mirkin, 2007; La Spada and Taylor, 2010; McMurray, 2010; DeJesus-Hernandez et al., 2011; Renton et al., 2011; Usdin et al., 2015; Haeusler et al., 2016). Most of these diseases are caused by unstable dynamic mutations that usually increase in size during meiotic divisions and have been associated with neurologic disorders (Virtaneva et al., 1997; Cummings and Zoghbi, 2000; Renton et al., 2011; Haeusler et al., 2016). However, poly-A repeats in *HOXD13*, the causative mechanism of a non-neurological condition, synpolydactyly, represents an exception, in which the most likely mechanism leading to increased poly-A tracts are errors in DNA replication (Muragaki et al., 1996; Warren, 1997; Brown and Brown, 2004).

We have shown that an increased number of repeats at 5'UTR of *EIF4A3* causes Richieri-Costa-Pereira syndrome (RCPS; OMIM #268305), a rare autosomal-recessive disorder affecting craniofacial and limb development, mainly described in Brazilian patients (Favaro et al., 2011, 2014; Bertola et al., 2017). RCPS individuals show a distinctive allelic pattern, determined not only by the larger number of repeats (>14 as compared to up to 12 repeats in controls), but also by the presence of a unique motif containing G instead of A nucleotide (the 'disease-associated CGCA-20nt motif') (Favaro et al., 2014). As the origin of the RCPS disease alleles remains unknown, characterizing the 5'UTR of *EIF4A3* in a populational level could give us clues on the mechanisms that originate the *EIF4A3* pathogenic alleles (e.g., meiotic instability or unequal crossing-over events), in addition to providing insights on the chance of RCPS arising in other populations.

We and others have shown that *EIF4A3* downregulation in cellular and animal models leads to defective neural crest cell migration/differentiation and neural stem cell apoptosis during embryonic development, paralleling RCPS cranioskeletal defects and microcephaly, respectively (Mao et al., 2016; Miller et al., 2017). However, the molecular mechanism responsible for *EIF4A3* downregulation remains entirely unknown.

Therefore, this work was undertaken to investigate the origin of the pathogenic alleles containing the disease-associated CGCA-20nt motif, as well as to evaluate the functional effects of the 5'UTR motifs on *EIF4A3* expression. Insights into the origin and effect of these complex alleles will contribute to a better understanding of regulatory features of 5'UTR regions and their role in craniofacial and neural development.

## MATERIALS AND METHODS

### Ethics Approval Statement

The protocol was approved by the Ethics Committee of Instituto de Biociências at Universidade de São Paulo, Brazil (accession

number 1.463.852). All individuals donated biological samples after providing signed informed consent.

### DNA Samples

To characterize *EIF4A3* 5'UTR, 380 DNA samples from unaffected individuals unrelated to RCPS families were selected from the biorepository of CEGH-CEL. For haplotype analysis, 13 additional samples were used, 12 are also from CEGH-CEL (four unaffected individuals without CGCA-20nt motifs and four with CGCA-20nt motifs; four Brazilian RCPS patients bearing different allelic structures) and one sample of a RCPS patient, from the United Kingdom, was sent for diagnosis purposes from North East Thames Genetics Service.

In order to evaluate the effects of the motifs on *EIF4A3* expression, six DNA samples carrying *EIF4A3* alleles with distinct number of motifs were selected from the biorepository of Centro de Estudos do Genoma Humano e Células Tronco (CEGH-CEL) and used for luciferase reporter assay.

For methylation assessment, we used DNA samples of RCPS patients ( $n = 6$ ; homozygous for the 16 repeats allele) from Hospital de Reabilitação de Anomalias Craniofaciais da Universidade de São Paulo (HRAC-USP) and unaffected individuals ( $n = 7$ ; homozygous for the eight repeats allele) from CEGH-CEL.

All samples were extracted from peripheral blood using the Gentra Systems Autopure LS (AutoGen) according to the manufacturer's protocol.

### *EIF4A3* 5'UTR Characterization and Haplotype Analysis

Sanger sequencing of *EIF4A3* 5'UTR (NM\_014740.3) and five flanking SNPs (rs11150824, rs2289534, rs3829612, rs10782008, and rs12943620) were performed using BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and the ABI 3730 DNA Analyzer (Applied Biosystems). Sequences were analyzed using Sequencher 5.1 (Gene Codes Corporation) and Mixed Sequences Reader (Chang et al., 2012) software. The last one allowed us to better discriminate the alleles. In order to validate our analysis and obtain more reliability in our data, we randomly selected 30 heterozygous samples with different alleles' structure to sequence each allele separately, using the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) followed by Sanger sequencing.

To facilitate results interpretation, we referred SNPs markers according to their genomic relative position on the annotated plus strand (GRCh38/hg38): SNP1 (rs11150824) – SNP2 (rs2289534) – *EIF4A3* – SNP3 (rs3829612) – SNP4 (rs10782008) – SNP5 (rs12943620). Primers were designed using Primer-Blast<sup>1</sup> (Untergasser et al., 2012) and are described in **Supplementary Table S1**. Linkage disequilibrium and haplotypes were inferred using Haploview software (Barrett et al., 2005).

### Luciferase Assay

Although the larger alleles (15 or 16 repeats) have been associated with decreased *EIF4A3* expression, the causal

<sup>1</sup><https://www.ncbi.nlm.nih.gov/tools/primer-blast/>

relationship between number and/or pattern of repeats and gene downregulation is still unknown. Therefore, we investigated the role of the 5'UTR motifs on *EIF4A3* expression by luciferase assay.

Sequences of interest of unaffected and RCPS individuals were amplified by PCR (primer sequences in **Supplementary Table S1**), purified using the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) and cloned into the pGL4.24[*luc2P*/minP] vector (Promega), upstream of a minimal promoter and the *luc2P* gene. Sanger sequencing confirmed all constructs: pGL4.24 vectors carrying control alleles with 4 repeats (3 CACA-20-nt and 1 CA-18-nt), 7 repeats (4 CACA-20-nt, 1 CA-18-nt, 1 CACA-20-nt, and 1 CA-18-nt), 10 repeats (7 CACA-20-nt, 1 CA-18-nt, 1 CACA-20-nt, and 1 CA-18-nt) and 12 repeats (1 CACA-20-nt, 10 CGCA-20nt, and 1 CA-18-nt); pGL4.24 vectors carrying pathogenic alleles with 14 repeats (2 CACA-20nt, 10 CGCA-20nt, 1 CACA-20nt, and 1 CA-18nt) and 16 repeats (1 CACA-20nt, 13 CGCA-20nt, 1 CACA-20nt, and 1 CA-18nt). In order to investigate the individual effect of each motif, differing in the composition of central nucleotides, we also constructed vectors carrying sequences with only 1 CA-18-nt, 1 CACA-20-nt, and 1 CGCA-20nt, synthesized by Integrated DNA Technologies (IDT).

Human embryonic kidney (HEK) 293T cells, cultured in high glucose DMEM supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum (FBS) (all provided by Life Technologies), were plated in 96-well plates 24 h prior to transfection ( $2 \times 10^4$  cells/well). Transient transfections were performed in triplicate by using TurboFectin 8.0 (OriGene) according to the manufacturer's instruction. Cells were cotransfected with 180 ng of the pGL4.24 constructs and 20 ng of the pRL-SV40 vector (Promega) containing the Renilla luciferase gene, used as a transfection control. The plasmid pLuc generated from pGL3-control template was used as positive control (Soltys et al., 2013). Forty-eight hours after DNA transfection, luciferase activity was measured with the Dual-Glo<sup>®</sup> Luciferase Assay System in a GloMax Multi 96-microplate Luminometer (Promega). Firefly luminescence results were normalized by Renilla luminescence and the relative luciferase activity was determined. Statistical analyses were performed with one-way ANOVA and Tukey *post hoc* test. Significance was set at  $p < 0.05$ .

## Methylation Assay

Indeed, methylated CpGs are involved in gene repression especially when occurring at promoter/5'UTR (Feil and Fraga, 2012; Schübeler, 2015). Since *EIF4A3* 5'UTR motifs are CG-rich and the disease CGCA-20nt allele shows increased number of CpGs compared to control alleles, we evaluated DNA hypermethylation as a plausible mechanism behind *EIF4A3* downregulation in RCPS patients.

One microgram of genomic DNA from each sample were submitted to bisulfite conversion using EpiTect Bisulfite Conversion Kit (QIAGEN). Bisulfite converted DNA was subsequently used for PCR, in which primers were designed

with MethPrimer<sup>2</sup> (Li and Dahiya, 2002) and are shown in **Supplementary Table S1**. Amplicons were checked by agarose electrophoresis and cloned using the TOPO TA Cloning Kit for sequencing (Thermo Fisher Scientific). Sanger sequencing was carried out for 10 clones per sample using the BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and the ABI 3730 DNA Analyzer (Applied Biosystems). Sequencing files were then analyzed for methylation quantification using the online tool BISMA – Bisulfite Sequencing DNA Methylation Analysis (Rohde et al., 2010) with lower threshold conversion rate at 95%, lower threshold sequence identity at 90%, upper threshold of N-sites at cytosine positions at 20% and per threshold gaps allowed at 20% as filtering parameters. Methylation values were computed and differences between groups tested using Fisher's Exact Test. Significance was set at  $p < 0.05$ .

## RESULTS

The *EIF4A3* 5'UTR is characterized by the presence of 18- or 20-nucleotide-long motifs differing in the composition of central nucleotides, namely CA-18nt (TCGGCAGCGG**C**AGCGAGG), CACA-20nt (TCGGCAGCGG**CAC**AGCGAGG), and CGCA-20nt (TCGGCAGCGG**CGC**AGCGAGG) (Favaro et al., 2014). Unaffected individuals have 3–12 repeats composed mostly by CA-18nt and CACA-20nt motifs, while RCPS patients have 14–16 repeats, with a higher number of the CGCA-20nt motif (Favaro et al., 2014). The motif CGCA-20nt will be herein referred as 'disease-associated CGCA-20nt motif.'

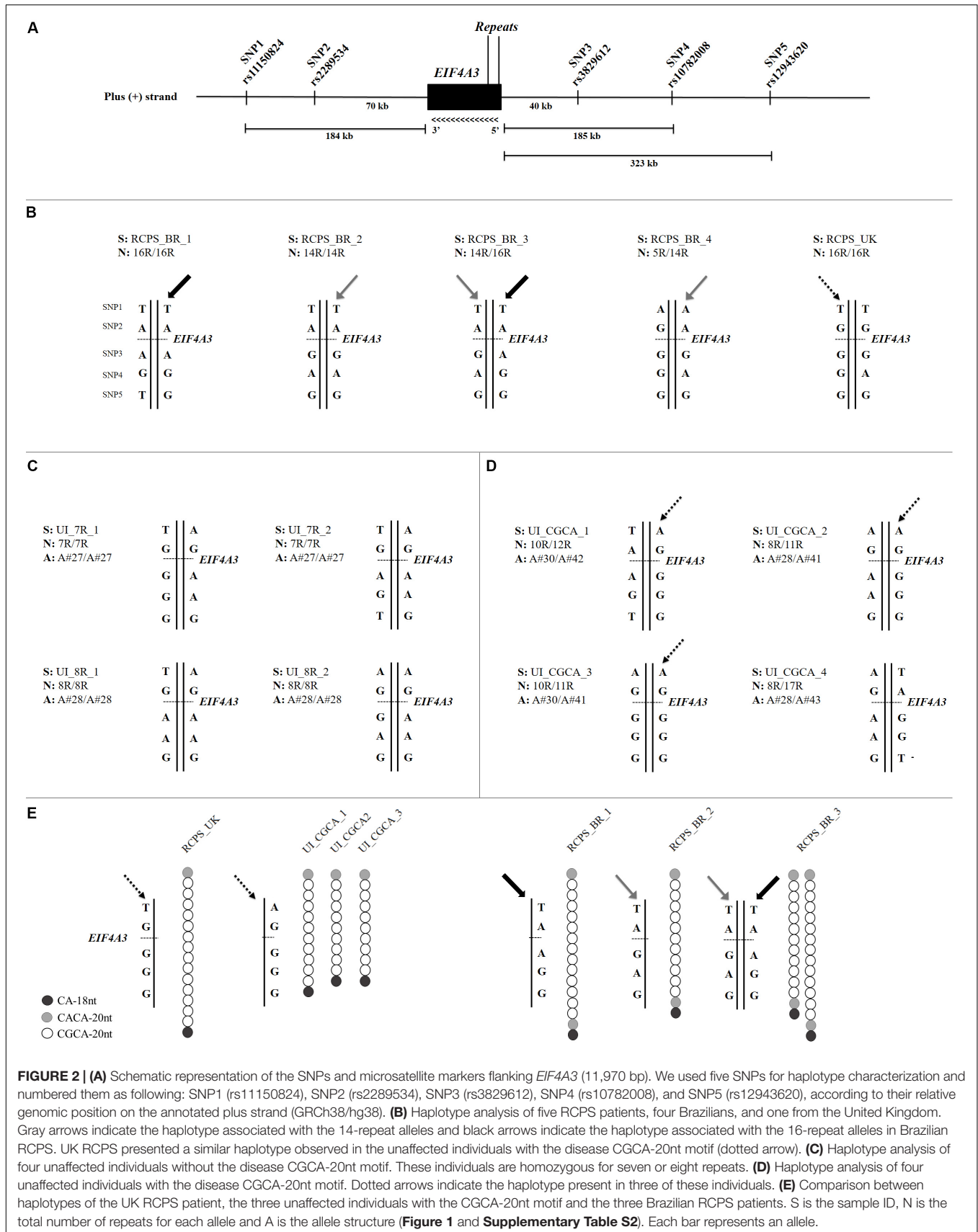
### Unaffected Individuals Are Mostly Heterozygous and May Also Present CGCA-20nt Motifs

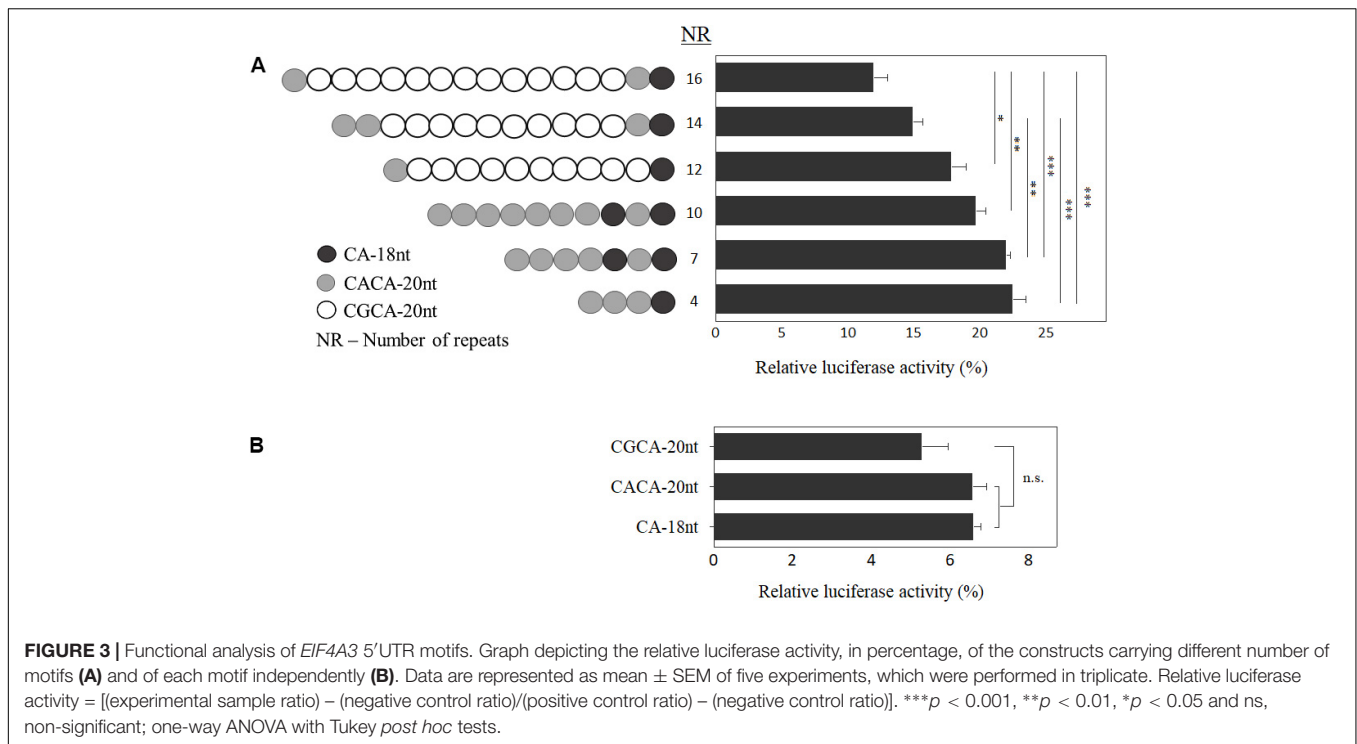
Sanger sequencing of *EIF4A3* 5'UTR in 380 unaffected individuals revealed 43 different alleles (**Figure 1**) in heterozygosis in 85% of individuals. The total number of repeats per allele varied from 2 to 17, and the most common alleles contained 7 (25.46%) or 8 (23.21%) repeats (**Supplementary Figure S1**). The number and organization of the CA-18nt or CACA-20nt motifs varied between the alleles. We identified not only alleles containing a single motif, but also complex alleles with multiple organizations of the CA-18nt or CACA-20nt motifs (**Figure 1**). For example, one allele exclusively constituted by the CA-18nt motif was found. As for the other alleles, the number of the first CACA-20nt motifs varied from 2 to 8 with or without different combinations of CA-18nt motifs (**Figure 1**). The most common allele contained seven repeats and was constituted by four repeats of CACA-20nt motif followed by 1 CA-18nt, 1 CACA-20nt, and 1 final CA-18nt. We also identified two atypical alleles comprising 12- or 20-nucleotide-long sequences inserted between motifs, which do not align to any viral sequences or neighboring genes (data not shown). Finally, the disease-associated CGCA-20nt motif, originally found only among RCPS patients, was identified here in eight heterozygous control

<sup>2</sup>[www.urogene.org/methprimer/](http://www.urogene.org/methprimer/)









between RCPS and controls, as both presented low methylation levels.

## DISCUSSION

Expansions at non-coding regions have been extensively described in neurological disorders, and characterization of these regions have greatly contributed to the understanding of novel regulatory mechanisms (Gatchel and Zoghbi, 2005; Mirkin, 2007; La Spada and Taylor, 2010; McMurray, 2010; Russo et al., 2015; Usdin et al., 2015; Haeusler et al., 2016). Despite the great advances in genome sequence analysis, DNA of repetitive regions is still difficult to be sequenced. In fact, the *EIF4A3* 5'UTR is not covered in GnomAD database, which reinforces the importance of characterizing this region through Sanger sequencing.

In this study, Sanger sequencing analysis of *EIF4A3* 5'UTR in 380 unaffected individuals revealed 43 different alleles, with the most common alleles containing seven or eight repeats. Some of these alleles presented only one type of motif (CA-18nt or CACA-20nt) with different total number of repeats, while others presented a visible combination of these two common motifs, suggesting that these alleles may have originated through unequal crossing-over events. These results show its polymorphic nature and confirm the structural complexity, and uniqueness of this region, which was not comparable to any gene in which dynamic pathogenic expansions at non-coding regions had been reported (Brook et al., 1992; Campuzano et al., 1996; Moseley et al., 2006; Daughters et al., 2009; Galloway and Nelson, 2009; Sato et al., 2009; DeJesus-Hernandez et al., 2011; Kobayashi et al., 2011). In this enlarged cohort, 1% of the

alleles (8/760) harbors the disease-associated CGCA-20nt motif. Interestingly, the largest alleles in this cohort ( $\geq 11$  repeats), including one with 17 repeats, contained the disease CGCA-20nt motif. These results suggest that RCPS could occur in any population containing alleles with the CGCA-20nt motif. Indeed, one of the patients here included is from United Kingdom (Bertola et al., 2017).

The haplotype analyses were performed in order to get insights on the origin of the pathogenic alleles with increased number of repeats. Results revealed that the pathogenic alleles with 14 and 16 repeats of Brazilian patients have distinct origins, which in turn are different from the haplotypes of the UK RCPS patient. These results suggest that the pathogenic alleles have arisen more than once. It is of note that the UK RCPS patient shares a common haplotype and also a similar motif structure with three unaffected individuals carrying the disease-associated CGCA-20nt motif, suggesting a common ancestral among them. It is possible that, similarly to the alleles in the control population, the affected alleles may have arisen also through unequal crossing. This hypothesis is also supported by the observation that the number of repeats at *EIF4A3* 5'UTR seems to be stable across generations (Favaro et al., 2014). This phenomenon is more comparable to the one observed in synpolydactyly, in which the poly-A at the 3' end of *HOXD13* might have originated by unequal crossing over as it is quite stable when transmitted across generations (Warren, 1997), as opposed to dynamic mutations observed in neurological conditions, which have arisen only once (Virtaneva et al., 1997; Gatchel and Zoghbi, 2005; Mirkin, 2007; La Spada and Taylor, 2010; McMurray, 2010; DeJesus-Hernandez et al., 2011; Renton et al., 2011; Usdin et al., 2015; Haeusler et al., 2016).

Next, we demonstrated that the number and allele sequence of motifs at 5'UTR is involved in *EIF4A3* expression. These results thus suggest a potential *cis*-acting regulatory mechanism for these motifs on gene expression and confirm our previous finding that the affected alleles were associated with *EIF4A3* downregulation in cells from different tissues from RCPS patients, including peripheral blood (Favaro et al., 2014; Miller et al., 2017). DNA hypermethylation is not the mechanism for *EIF4A3* downregulation in RCPS patients, as RCPS and control blood samples showed similar methylation levels. Hypomethylation, as observed at *EIF4A3* 5'UTR, is consistent with transcriptional activity (Baubec and Schübeler, 2014; Schübeler, 2015). There could be alternative mechanisms by which the increased number of motifs could repress *EIF4A3* expression: the disease CGCA-20nt could create binding sites for repressor proteins and act as a silencer; or more complex mechanisms, including post-translational events, could be involved. Further functional studies are needed to pinpoint the exact molecular mechanism underlying *EIF4A3* downregulation.

The inverse correlation between number of repeats and *EIF4A3* expression levels suggests that the structure of the 5'UTR *EIF4A3* may modulate the phenotype. Indeed, a broader phenotypic spectrum within RCPS has been observed, ranging from individuals with severe phenotype (homozygous for the 16-repeat allele or heterozygous for the 15- and 16-repeat alleles) to individuals with less severe skeletal involvements, harboring smaller number of motifs (homozygous for the 14-repeat allele or compound heterozygous for 14 repeats and a point mutation (c.809A>G) (Favaro et al., 2014; Bertola et al., 2017). Despite limitations in sample size, these results suggest that RCPS phenotypic variability depends upon the number and allele sequence of repeats at the 5'UTR of *EIF4A3*.

## CONCLUSION

In summary, we provide evidence supporting that the *EIF4A3* 5'UTR is highly polymorphic, comprising at least 43 different alleles, which may have originated through extensive recombination in this region. Haplotype analysis in control and in affected individuals suggests that there are more than one haplotype associated with the disease and RCPS alleles might have originated from unequal crossing-over events. We also provided further evidence that *EIF4A3* 5'UTR is a regulatory region, and that *EIF4A3* downregulation in RCPS is not mediated by CpG methylation. Moreover, our findings provide insights to explain clinical variability in RCPS.

## REFERENCES

- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Baubec, T., and Schübeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. *Curr. Opin. Genet. Dev.* 25, 85–92. doi: 10.1016/j.gde.2013.11.015

## AUTHOR CONTRIBUTIONS

DB, RP, and MZ provided the Brazilian DNA samples. AG and EW provided DNA sample and clinical information from the UK RCPS patient. GH, CM, and MP-B conceived and designed the study. GH performed the *EIF4A3* 5'UTR characterization and haplotype analyses. LB assisted with haplotype and linkage disequilibrium analyses. CM carried out the luciferase assay and LA performed the methylation assay. GH, CM, LA, LB, GK, and MP-B discussed and interpreted the results and wrote the manuscript. GH and CM have a major contribution in writing the manuscript. All authors contributed to the final version and approved the manuscript.

## FUNDING

This work was funded by Centro de Pesquisa, Inovação e Difusão (CEPID, 2013/08028-1), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2015/21781-6), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, PQ10/2011), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, 1470032).

## ACKNOWLEDGMENTS

We thank Dr. Meire Aguea and Dr. Antonia Cerqueira (IB-USP) for technical support. We also thank Dr. Marcelo Nobrega and Dr. Débora Sobreira (University of Chicago, United States) who donated the sequences for constructs used in the luciferase assay.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00149/full#supplementary-material>

**FIGURE S1** | Distribution of number of alleles per total number of repeats identified in the 380 samples of unaffected individuals. This analysis did not take in consideration the motifs' structure.

**TABLE S1** | Primer sequences.

**TABLE S2** | List of the 13 samples used in haplotype analysis and the description of motifs' structure reported in each allele.

**TABLE S3** | Methylation levels in the analyzed regions in RCPS patients and control individuals;  $p = 0.0461$ , Fisher's Exact Test.

- Bertola, D. R., Hsia, G., Alvizi, L., Gardham, A., Wakeling, E. L., Yamamoto, G. L., et al. (2017). Richieri-Costa-Pereira syndrome: expanding its phenotypic and genotypic spectrum. *Clin. Genet.* 93, 800–811. doi: 10.1111/cge.13169
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68, 799–808. doi: 10.1016/0092-8674(92)90154-5

- Brown, L. Y., and Brown, S. A. (2004). Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* 20, 51–58. doi: 10.1016/j.tig.2003.11.002
- Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423–1427. doi: 10.1126/science.271.5254.1423
- Chang, C.-T., Tsai, C.-N., Tang, C. Y., Chen, C.-H., Lian, J.-H., Hu, C.-Y., et al. (2012). Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. *ScientificWorldJournal* 2012:365104. doi: 10.1100/2012/365104
- Cummings, C. J., and Zoghbi, H. Y. (2000). Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* 9, 909–916. doi: 10.1093/hmg/9.6.909
- Daughters, R. S., Tuttle, D. L., Gao, W., Ikeda, Y., Moseley, M. L., Ebner, T. J., et al. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet.* 5:e1000600. doi: 10.1371/journal.pgen.1000600
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256. doi: 10.1016/j.neuron.2011.09.011
- Favaro, F. P., Alvizi, L., Zechi-Ceide, R. M., Bertola, D., Felix, T. M., de Souza, J., et al. (2014). A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *Am. J. Hum. Genet.* 94, 120–128. doi: 10.1016/j.ajhg.2013.11.020
- Favaro, F. P., Zechi-Ceide, R. M., Alvarez, C. W., Maximino, L. P., Antunes, L. F., Richieri-Costa, A., et al. (2011). Richieri-Costa-Pereira syndrome: a unique acrofacial dysostosis type. An overview of the Brazilian cases. *Am. J. Med. Genet. A* 155A, 322–331. doi: 10.1002/ajmg.a.33806
- Feil, R., and Fraga, M. F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97–109. doi: 10.1038/nrg3142
- Galloway, J. N., and Nelson, D. L. (2009). Evidence for RNA-mediated toxicity in the fragile X-associated tremor/ataxia syndrome. *Future Neurol.* 4:785. doi: 10.2217/fnl.09.44
- Gatchel, J. R., and Zoghbi, H. Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* 6, 743–755. doi: 10.1038/nrg1691
- Haeusler, A. R., Donnelly, C. J., and Rothstein, J. D. (2016). The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. *Nat. Rev. Neurosci.* 17, 383–395. doi: 10.1038/nrn.2016.38
- Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., et al. (2011). Expansion of intronic GGCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am. J. Hum. Genet.* 89, 121–130. doi: 10.1016/j.ajhg.2011.05.015
- La Spada, A. R., and Taylor, J. P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* 11, 247–258. doi: 10.1038/nrg2748
- Li, L.-C., and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18, 1427–1431. doi: 10.1093/bioinformatics/18.11.1427
- Mao, H., McMahon, J. J., Tsai, Y. H., Wang, Z., and Silver, D. L. (2016). Haploinsufficiency for core exon junction complex components disrupts embryonic neurogenesis and causes p53-mediated microcephaly. *PLoS Genet.* 12:e1006282. doi: 10.1371/journal.pgen.1006282
- McMurray, C. T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* 11, 786–799. doi: 10.1038/nrg2828
- Miller, E. E., Kobayashi, G. S., Musso, C. M., Allen, M., Ishiy, F. A. A., de Caires, L. C., et al. (2017). EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Hum. Mol. Genet.* 26, 2177–2191. doi: 10.1093/hmg/ddx078
- Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature* 447, 932–940. doi: 10.1038/nature05977
- Moseley, M. L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A. K., Daughters, R. S., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.* 38, 758–769. doi: 10.1038/ng1827
- Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 272, 548–551. doi: 10.1126/science.272.5261.548
- Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268. doi: 10.1016/j.neuron.2011.09.010
- Rohde, C., Zhang, Y., Reinhardt, R., and Jeltsch, A. (2010). BISMAs—fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* 11:230. doi: 10.1186/1471-2105-11-230
- Russo, A., Pacchierotti, F., Cimini, D., Ganem, N. J., Genescà, A., Natarajan, A. T., et al. (2015). Genomic instability: crossing pathways at the origin of structural and numerical chromosome changes. *Environ. Mol. Mutagen.* 56, 563–580. doi: 10.1002/em.21945
- Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., et al. (2009). Spinocerebellar ataxia type 31 is associated with “inserted” pentanucleotide repeats containing (TGGA)n. *Am. J. Hum. Genet.* 85, 544–557. doi: 10.1016/j.ajhg.2009.09.019
- Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326. doi: 10.1038/nature14192
- Soltys, D. T., Rocha, C. R., Lerner, L. K., de Souza, T. A., Munford, V., Cabral, F., et al. (2013). Novel XPG (ERCC5) mutations affect DNA repair and cell survival after ultraviolet but not oxidative stress. *Hum. Mutat.* 34, 481–489. doi: 10.1002/humu.22259
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596
- Usdin, K., House, N. C., and Freudenreich, C. H. (2015). Repeat instability during DNA repair: insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* 50, 142–167. doi: 10.3109/10409238.2014.999192
- Virtaneva, K., D'Amato, E., Miao, J., Koskiniemi, M., Norio, R., Avanzini, G., et al. (1997). Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nat. Genet.* 15, 393–396. doi: 10.1038/ng0497-393
- Warren, S. T. (1997). Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science* 275, 408–409. doi: 10.1126/science.275.5298.408

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hsia, Musso, Alvizi, Brito, Kobayashi, Pavanello, Zatz, Gardham, Wakeling, Zechi-Ceide, Bertola and Passos-Bueno. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.